



Statistics for Business: Formulas

- **Sample Variance, Covariance and Correlation**

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n-1}; \quad s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{n-1};$$

$$r = \frac{s_{xy}}{\sqrt{s_x^2} \sqrt{s_y^2}} = \frac{s_{xy}}{s_x s_y}$$

- **Combinatorics:**

$$n! = n(n-1)(n-2)\dots(2)(1); \quad P_r^n = \frac{n!}{(n-r)!}; \quad C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

- **Probability**

- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$; $\Pr(A \cap B) = \Pr(B) \Pr(A|B) = \Pr(A) \Pr(B|A)$
- Let E_1, E_2, \dots, E_k be a finite partition of the sample space S and $A \subset S$. Then

$$\Pr(A) = \Pr(A \cap E_1) + \Pr(A \cap E_2) + \dots + \Pr(A \cap E_k).$$

- **Variance, Covariance and Correlation**

$$\text{Var}(X) = E[(X - \mu)^2] = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2;$$

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - E(X)E(Y) = E(XY) - \mu_X \mu_Y;$$

$$\rho = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}}$$

$$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y);$$

$$\text{Cov}(a + bX, c + dY) = bd \text{Cov}(X, Y)$$

- **Discrete Random Variables**

$$\text{Bernoulli} : P(x) = p^x (1-p)^{1-x}; \quad x = 0, 1$$

$$\text{Binomial} : P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}; \quad x = 0, 1, 2, \dots, n$$

- **Sampling Distributions – Confidence Intervals – Hypothesis tests**

- If $X \sim N(\mu, \sigma^2)$, and $\{X_1, \dots, X_n\}$ is a random sample, then

$$\text{if } \sigma^2 \text{ is known or } n \text{ is large} : \frac{\bar{X} - \mu}{\sigma_X / \sqrt{n}} \sim N(0, 1) \text{ or } \frac{\bar{X} - \mu}{S_X / \sqrt{n}} \rightarrow N(0, 1)$$

$$\text{if } \sigma^2 \text{ is unknown and } n \text{ is small} : \frac{\bar{X} - \mu}{S_x / \sqrt{n}} \sim t_{n-1}$$

- If $\{X_1, \dots, X_n\}$ and $\{Y_1, \dots, Y_n\}$ are dependent (paired or matched) random samples and n is small, then with $d_i = x_i - y_i$;

$$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} \sim t_{n-1}; \text{ where } \bar{d} = \frac{1}{n} \sum_{i=1}^n d_i; \text{ and } s_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2}$$

- If $X \sim N(\mu_x, \sigma_x^2), Y \sim N(\mu_y, \sigma_y^2)$ and $\{X_1, \dots, X_n\}$ and $\{Y_1, \dots, Y_n\}$ are independent random samples, then

$$\text{if } \sigma_x^2 \text{ and } \sigma_y^2 \text{ known} : \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{(\sigma_x^2/n_x) + (\sigma_y^2/n_y)}} \sim N(0, 1);$$

$$\text{if } \sigma_x^2 \text{ and } \sigma_y^2 \text{ unknown but } n_x, n_y \text{ are large} : \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{(s_x^2/n_x) + (s_y^2/n_y)}} \rightarrow N(0, 1);$$

$$\text{if } \sigma_x^2 \text{ and } \sigma_y^2 \text{ unknown, but equal} : \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{(s_p^2/n_x) + (s_p^2/n_y)}} \sim t_{n_x+n_y-2};$$

$$\text{where } s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$$

- $\hat{\theta} \pm c \cdot \text{SE}(\hat{\theta})$ [Point Estimate \pm (Reliability Factor)·(Standard Error)]
- $TS = (\hat{\theta} - \theta_0)/\text{SE}(\hat{\theta})$

• Correlation

- The test statistic for testing if $\rho = 0$ is:

$$t = r\sqrt{\frac{n-2}{1-r^2}} \sim t_{n-2}.$$

• Simple Regression Prediction Intervals

- Confidence interval estimate for the *expected value* of y given a particular x_0

$$\hat{y}_0 \pm t_{n-2, \alpha/2} \cdot s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} = (b_0 + b_1 x_0) \pm t_{n-2, \alpha/2} \cdot s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

- Confidence interval for an *actual observed value* of y given a particular x_0

$$\hat{y}_0 \pm t_{n-2, \alpha/2} \cdot s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} = (b_0 + b_1 x_0) \pm t_{n-2, \alpha/2} \cdot s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

$$\text{where } s_e = \sum_{i=1}^n e_i^2 / (n - 2) = SSE / (n - 2).$$